ARMY RESEARCH LABORATORY

The Effects of Movement and Intravehicular Versus Intervehicular Communication on C²V Crew Performance: Limited User Test Phase III

Hall P. Beck
Linda G. Pierce

ARL-TR-1428

JANUARY 1998

prepared by

Appalachian State University
Department of Psychology
Boone, North Carolina 28608

under contract

DAAL03-91-C-0034

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5425

ARL-TR-1428 January 1998

The Effects of Movement and Intravehicular Versus Intervehicular Communication on C2V Crew Performance: Limited User Test Phase III

Hall P. Beck
Appalachian State University

Linda G. Pierce
Human Research & Engineering Directorate

prepared by

Appalachian State University
Department of Psychology
Boone, North Carolina 28608

under contract

DAAL03-91-C-0034

Abstract

The procedure and results of the group performance component of the command and control vehicle (C^2V) Limited User Test (LUT) Phase III are described in this report. The test was conducted to examine (a) the effects of movement on the ability of crews to work effectively as a team, (b) terrain impacts on team performance tasks, and (c) the effect of distributed team operations. Sixteen National Guardsmen, divided into four-person teams, served as participants. The evaluation design was similar to a 2 (Movement: Stationary, Moving) x 2 (Terrain: Paved, Course A) x 2 (Communication: Intravehicle, Intervehicle) with the baseline occupying the position of the nonfitting control arrangement. The effects of movement on team performance were evaluated by conducting some trials while the C^2V was stationary and other trials while it was moving. The influence of terrain on team performance was studied by conducting some trials on Course A of the Perryman test course and the remaining trials on a paved 3-mile course. In the intravehicular communication condition, the four members of a team were housed in the same C^2V and worked together on the same task. Teammates had visual contact and communicated verbally via intercom. Two teammates were in each C^2V for the intervehicular manipulation. It was concluded that the C^2V environment impaired all group performance tasks, especially those that appeared to demand a great degree of coordination and integration. Team performance was below the baseline when crews were housed in the C^2V, regardless of whether the vehicle was stationary or moving, although movement increased the deleterious impact of the C^2V on group performance. The impact of terrain on performance was inconclusive, possibly because of the small sample size and the limited number of situational conditions examined. If the C^2V is to become a prominent part of the 21st century Army's arsenal, then additional experimentation must be conducted to assess implications for team performance during a variety of conditions using validated task procedures.

ACKNOWLEDGMENTS

This work was supported by the Human Research and Engineering Directorate under the auspices of the U.S. Army Research Office Scientific Services Program administered by Battelle (Delivery Order 1853, Contract No. DAAL03-91-C-0034).

CONTENTS

TABLES

EXECUTIVE SUMMARY

The Army is developing a tracked command and control vehicle (C^2V) with a speed comparable to that of the combatant force. This report describes the procedure and results of the group performance component of the Limited User Test (LUT) Phase III. The main objectives of the LUT III were to (a) discover if movement impaired the ability of crews to work effectively as a team, (b) determine if performance deteriorated when soldiers in adjacent C^2Vs were required to integrate their activities, and (c) ascertain the impact of terrain on group performance tasks.

The LUT III used two C^2V prototypes manufactured by United Defense Industries. Sixteen National Guardsmen (NG), divided into four-person teams, served as participants. The Guardsmen manned four workstations in the vehicle's mission module. The evaluation design was similar to a 2 (Movement: Stationary, Moving) x 2 (Terrain: Paved, Course A) x 2 (Communication: Intravehicle, Intervehicle) with the baseline occupying the position of the nonfitting control arrangement. Four group performance tasks yielded 10 dependent variables. All tests were conducted at Aberdeen Proving Ground, Maryland.

The effects of movement on group performance were evaluated by conducting some trials while the C^2V was stationary and other trials while it was moving. For safety reasons, the vehicle's top speed was restricted to 20 miles per hour. The influence of terrain on group performance was studied by conducting some trials on Course A of the Perryman Track and the remaining trials on a paved 3-mile course.

In the intravehicular communication condition, the four members of a team were housed in the same C^2V and worked together on the same task. Teammates had visual contact and communicated verbally via intercom. Two teammates were in each C^2V for the intervehicular manipulation. The single channel ground airborne system (SINCGARS) was used to communicate between C^2Vs. Baseline was a benign condition, in which participants worked in a quiet temperature-controlled room

The principal findings of the LUT III group performance tests were

1. Crews working in C^2Vs did not perform as well as teams working under baseline conditions. The overall performance of teams in stationary C^2Vs was 13% below baseline.

2. Vehicle movement augmented the deleterious effects of the C^2V environment on team performance. Housing crews in moving vehicles produced a 22% decline in performance below baseline.

3. The C^2V environment impaired all tasks. The C^2V had its most adverse impact on tasks that required the greatest integration of teammates' activities.

4. During three of four tasks, performance was better on Course A than on the paved 3-mile course. The small sample size suggests caution in making any conclusions regarding the effects of terrain on group performance.

5. The results of the LUT III group performance tests are most applicable to situations in which crews are not required to process information at a rapid rate. The findings are also pertinent to vehicles moving at slow to moderate speeds.

If the C^2V is to become a prominent part of the 21st century Army's arsenal, then it should be developed so as to maximize group task performance as an analog to command staff performance. Team performance should be assessed in more technologically advanced C^2Vs than the prototypes used in LUT III. Also, evaluations must be on a larger scale so that the interactions between the variables that control collective behavior can be ascertained.

THE EFFECTS OF MOVEMENT AND INTRAVEHICULAR VERSUS INTERVEHICULAR COMMUNICATION ON C^2V CREW PERFORMANCE: LIMITED USER TEST PHASE III

INTRODUCTION

During Operation Desert Storm, commanders were often unable to keep pace with their forces. To rectify this problem, the Army is developing a tracked command and control vehicle (C^2V) with a speed comparable to that of the combatant force. The C^2V will replace the less mobile M-577, which entered service in 1963. The new vehicle will be a highly automated command post, able to communicate horizontally and vertically via a complex network of sensors and data links. Commanders and their staffs will receive a "real time" common picture of the battlefield, enabling exact and prompt direction of forces.

The advanced technologies that make the C^2V possible create a set of serious information processing problems. Sophisticated sensors will inundate the C^2V with vast quantities of data, augmenting the likelihood of information overload. New and highly efficient group interaction patterns must be developed if C^2V crews are to successfully manage rapid rates of information input. Space limitations within the C^2V will also emphasize the importance of teamwork. Current configurations allow for only four workstations. Precise coordination and automation must offset the liability of small crew size if the advantages of increased mobility and enhanced information sensitivity are to be fully realized.

Initial tests of the C^2V primarily evaluated equipment and individual task performance. Although these preliminary tests yielded useful data, apparati and individual skills are only pertinent in that they contribute to collective performance. Many command and control (C2) tasks require synchronization, the ability of crews to coordinate their activities and to achieve a unison of action. The ultimate value of the C^2V will be determined by whether it facilitates or impairs team performance. The Limited User Test (LUT) Phase III test was a notable advancement over preceding evaluations of the C^2V in that it included a series of group performance tasks.

This report describes the procedure and results of the group performance component of the LUT III. A main objective of the LUT III was to discover if movement impaired the ability of crews to work as a team. A second important question was whether performance deteriorated when the task required soldiers in adjacent C^2Vs to integrate their activities. The effects of between- versus within-vehicle communication were examined by housing team members in different C^2Vs during some trials and putting the entire crew in the same vehicle during other trials. In addition, crews were tested on different courses to assess the impact of terrain on crew performance.

METHOD

Design

This evaluation did not precisely conform to any experimental or quasi-experimental design. It is similar to an approach that Himmelfarb (1975) suggested using when the control arrangement does not mesh with the factorial design. If a Himmelfarb type of structure is imposed, the design may be viewed as a 2 (Movement: Stationary, Moving) x 2 (Communication Type: Intervehicle, Intravehicle) x 2 (Terrain: Paved, Course A) with the baseline occupying the position of the nonfitting control arrangement. The design was multivariate; ten dependent variables generated from four group performance tasks.

Participants

Sixteen Pennsylvania National Guardsmen (NG) (all male) were selected as test players. Potential participants were briefed at their home stations about the experimental procedures and risks involved in the evaluation. They also completed a short survey to determine their susceptibility to motion sickness and to ensure their familiarity with tracked vehicle operations. All players volunteered to participate in the evaluation as a special duty assignment and completed an informed consent form before arriving at the test site.

Participants were divided into two equal sections. Sections were composed of two, four-person teams. The senior ranking individual in each section served as the section chief and the senior person on each team was the team chief. Responsibilities of the team chiefs included reporting personnel and equipment status to the section chief. Section chiefs transmitted the status of each team to test personnel daily.

Apparatus

The C^2V

The LUT III used two C^2V prototypes, manufactured by United Defense Industries. Each prototype was equipped with an environmental heating and cooling system and nuclear, biological, and chemical (NBC) protection. The C^2V uses a chassis that is similar to that of the multiple launch rocket system and is divided into a cab section and a mission module. The cab contains seats and equipment for a driver and a track commander. The objectives of LUT III did not include assessments of either the driver or the track commander. Investigative personnel drove the vehicle and no tests involved the track commander.

The NG participants manned four workstations in the mission module. Three of the workstations faced to the side and the fourth workstation faced to the rear of the mission module. Each workstation included an adjustable seat that compensated for the individual's height and weight. When fielded, C^2V workstations will employ Army Tactical Command and Control System (ATCCS) equipment. However, ATCCS technology was not available for this evaluation. During LUT III, communication between crew members in the same vehicle was via intercom. The single channel ground airborne system (SINCGARS) was used to communicate with teammates in the other C^2V. A report by Martin Marietta Energy Systems (1993) provides a more detailed description of the equipment composing the C^2V.

Task Selection

A major issue in any investigation of collective behavior is deciding which tasks to include in the study. This was a particularly difficult problem in evaluating the C^2V because crews must perform a variety of tasks. One approach is to use tasks that C^2V crews will conduct in the field. Such a study would yield some useful findings, but the data would be of limited generality. For instance, determining the effects of movement on message transmission would reveal nothing about the impact of movement on the ability of commanders to develop battle plans. The conceptual challenge to evaluators is to decrease the number of potential investigative tasks without significantly reducing the generalizability of the findings.

A strategy for handling this problem is to derive a taxonomy of group functions that encompasses the realm of tasks that actual C^2V teams will perform. Presumably, teams will have fewer functions than tasks. After a useful taxonomy has been identified, laboratory tasks that are exemplars of those functions can be selected for testing.

This is the approach that Richard McGlynn and his associates used to select the group performance tasks for LUT III (McGlynn, Sutton, Demski, Sprague, & Pierce, in press). First, they developed a set of team functions based on a taxonomy proposed by Fleishman and Zaccaro (1992). One hundred fifty-two laboratory tasks were then reviewed and related to team functions. Each of these tasks was taken from the social or organizational psychology literature and had been shown in prior studies to be sensitive to environmental and group variables.

Potential tasks for the LUT III were evaluated according to their feasibility for administration in the C^2V and the likelihood that they tapped one and only one group function. Table 1 shows the tasks that McGlynn et al. selected for the LUT III and their associated functions. More detailed descriptions of particular tasks are deferred until the results section of this report.

Table 1

Tasks and Associated Team Functions

Task	Function
Sentence construction	Coordination
Social judgment	Error checking
Scrabble 2	Coordination
Quiz	Resource matching

Procedure

U.S. Army Research Laboratory (ARL) personnel administered the team performance tasks at Aberdeen Proving Ground, Maryland. Rick Tauson served as principal investigator; Bill Doss and Debbie Patton were co-investigators. The baseline condition was conducted in an environment designed to maximize group performance. Teams were tested in an amply lighted and temperature-controlled room. Crew members sat at tables, had visual contact with their teammates, and could easily hold discussions when the task permitted.

Before testing, participants were given an overview of the C^2V and the objectives of the evaluation. As part of this introduction, the equipment in the mission module was demonstrated. All trials in which teams worked in the C^2V were administered at the Perryman test course. One goal of the evaluation was to establish the effects of movement on group performance. Therefore, the C^2V was stationary during some trials and moving during others.

The influence of terrain on group performance was studied by varying the track on which the crew was tested. When the C^2V was moving, approximately half the trials were conducted on Perryman's Course A and the remainder on the paved 3-mile course. During LUT III, the C^2V was restricted to a top speed of 20 mph.

In the intravehicular communication condition, the four team members were housed in the same C^2V and worked together on the same task. Teammates had visual contact and communicated verbally via intercom. Intercoms were programmed before each test and excluded transmissions from the other C^2V. Two teammates were in each C^2V for the intervehicular manipulation. SINCGARS provided communication between vehicles.

RESULTS

In science, as in art, beauty is often found in simplicity. Given that the small sample size of this investigation restricted the use of inferential statistics, a very straightforward analysis appeared in order. The initial plan of analysis was to compute data across teams for each task and to compare the mean performances resulting from the independent variables. Facilitation or debilitation could be assessed by subtracting the mean baseline performance from the means of the experimental conditions.

Regrettably, the group performance component of LUT III contained violations of internal validity, negating the possibility of a series of straightforward mean comparisons. The only appropriate action is to bring internal validity issues to the forefront, taking the limitations that they impose upon data interpretations into consideration. The threats to internal validity in this evaluation were of two sorts. Some were design violations, affecting all group tasks. Other data collection problems were test specific. Breaches of internal validity attributable to design problems will be examined first, leaving specific data collection problems to be considered with the results of each test.

Carry-over Effects

Whenever a team is repeatedly assessed on the same or similar tasks, the potential exists for performance to change because of practice or carryover. Carry-over effects are an important area of inquiry in their own right but often create interpretive problems when performances are compared across trials. To illustrate, assume that the third time that a team worked on the Scrabble No. 2 Task they were communicating between vehicles and obtained a score of 65 points. The eighth time they worked on Scrabble No. 2, the team communicated within vehicles, scoring 85 points. Was the 20-point difference between trials the result of the conditions of interest (intervehicular versus intravehicular communication), carry-over effects, or both?

Carry-over effects are usually controlled by counterbalancing or treating trials as an independent variable (e.g., Christensen, 1980; Edwards, 1968). Scheduling and vehicle equipment problems prevented complete counterbalancing, and the small sample size precluded entering another independent variable into the analysis. The carry-over confound is a major problem, jeopardizing the integrity of the data. Unless the effects of carryover can be largely separated from the effects of the independent variables, the findings will be ambiguous and the conclusions of this evaluation will not meet minimal standards of scientific validity.

The gravity of the carry-over confound requires an effort to reduce its influence on the test results. The gist of the following strategy for controlling carryover was adopted from the behavioral sciences literature. Commonly used control procedures were combined and modified so that they could be applied to the LUT III group performance data. The method is presented in a step-by-step fashion, enabling the reader to decide how successfully independent variables have been distinguished from the effects of carryover. To elucidate the control procedure, Team A's Sentence Construction data are analyzed. Several criticisms of this approach are then discussed.

1. Repetitions of the task will be sequenced as trials. Team A was tested on parallel forms of the Sentence Construction Task nine times (see Table 2 and Figure 1). Baselines occurred on the first, fourth, and seventh trials of the sequence.

Table 2

Actual and Predicted Number of Words Formed by Team A on the Sentence Construction Task

	Trials								
	1	2	3	4	5	6	7	8	9
Scores	B	M-W-P	S-W-P	B	M-W-P	S-W-P	B	M-B-A	S-B-A
Actual	25.00	22.00	28.00	29.00	29.00	27.00	34.00	26.00	31.00
Predicted[a]	24.83	26.33	27.83	29.33	30.83	32.33	33.83	35.33	36.83
Difference[bc]	0.17	-4.33	0.17	-0.33	-1.83	-5.33	0.17	-9.33	-5.83

Note. B=baseline; Vehicle - S=stationary, M=moving; Communication - B=between vehicle, W=within vehicle; Course - A=Course A, P=paved or 3-mile course.
For example, 'M-W-P' represents a C^2V moving, using within-vehicle communication, on the paved course.
[a]Predicted score is based on the best fitting line calculated from the baseline trials.
[b]Difference score is the actual score minus the predicted score.
[c]A negative difference score indicated a performance inferior to baseline. A positive difference score shows a performance superior to baseline.

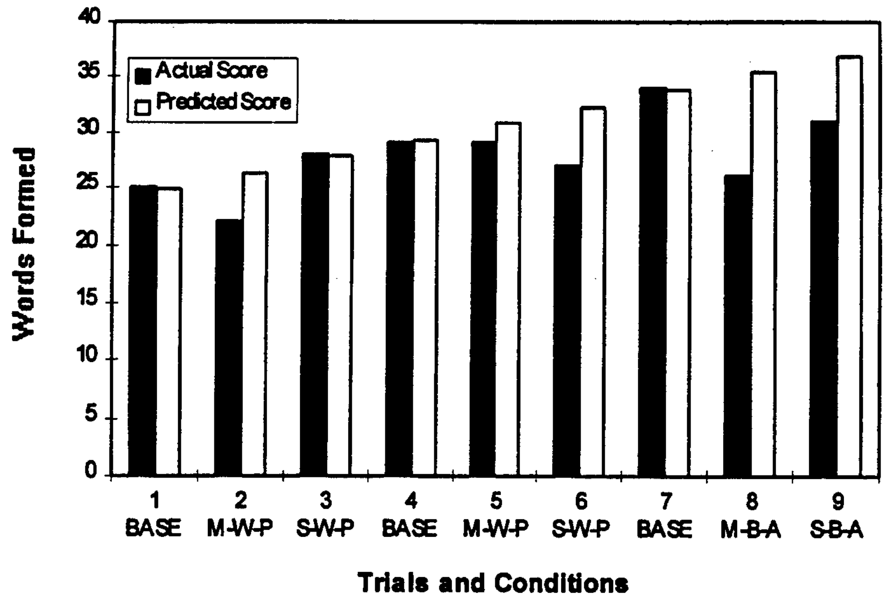

Figure 1. Words formed by Team A in the sentence construction task.

2. Performance of the C^2V crews during the LUT III may be considered the product of the experimental manipulations, carry-over effects, and random error. Random error includes the effects of all variables (e.g., abilities of crew members) other than the experimental conditions and carryover.

$$\text{Performance} = \text{Experimental Conditions} + \text{Carryover} + \text{Error}$$

Because the same experimental condition (baseline) was used on all trials, differences in performance are attributable to either carry-over effects or random error. For example, the difference in the performance of Team A, when the baseline was assessed on Trials 4 and 7, would be

$$\text{Baseline}_{(7)} - \text{Baseline}_{(4)} = [\text{Carryover}_{(7)} + \text{Error}_{(7)}] - [\text{Carryover}_{(4)} + \text{Error}_{(4)}]$$

3. Although the error for any given trial cannot be precisely determined, it can be estimated. By definition, random error is equally likely to increase or decrease performance on a particular trial. Statistically, the mean effect of all randomly distributed errors on performance is zero. Thus, zero is the best estimate of random error on any given trial. If error is assumed to be zero, our model shows that any performance difference between Trials 4 and 7 was attributable to carryover.

$$\text{Baseline}_{(7)} - \text{Baseline}_{(4)} = [\text{Carryover}_{(7)} + 0] - [\text{Carryover}_{(4)} + 0]$$

$$\text{Baseline}_{(7)} - \text{Baseline}_{(4)} = \text{Carryover}_{(7)} - \text{Carryover}_{(4)}$$

4. A Pearson r and a best fitting line were calculated from Team A's baseline data (Trials 1, 4, and 7). Team A formed 25, 29, and 34 words on Trials 1,4, and 7, respectively. Therefore, the data points (see Table 2) used in calculating the correlation were 1,25; 4,29; 7,34. The linear equation describing the Sentence Construction baseline performance of Team A was

$$\text{Words Formed} = 23.33 + 1.50\ (\text{Trial})$$

5. The best fitting line for Team A is an accelerating function, suggesting that carry-over effects are producing an increase in the dependent variable over trials (see "Predicted Scores" in Table 2 and Figure 1). The slope will be "0" and the best fitting line will be horizontal when carry-over effects are not present. A decelerating function implies that carry-over effects caused a reduction in the dependent measures.

Recall that this equation was calculated using baseline Trials 1, 4, and 7. Trial 2 was not a baseline condition. However, the linear equation can yield an estimate of what Team A's performance would have been if Trial 2 were conducted as baseline. Simply insert the trial number into the equation and compute.

$$\begin{aligned} \text{Words Formed} &= 23.33 + 1.50\ (\text{Trial}) \\ &= 23.33 + 1.50\ (2) \\ &= 26.33 \end{aligned}$$

The linear equation estimates that Team A would have formed 26.33 words if the second trial were baseline. Similar estimations were made for all trials (see Table 2 and Figure 1).

6. Team A formed 22 words on Trial 2. A difference score of -4.33 was obtained by subtracting the predicted baseline score from the actual score for that trial. Table 2 shows the experimental conditions and difference scores for Team A on each trial.

Difference scores provide a comparison of the experimental condition to the baseline after allowing for carry-over effects. If the difference score is a minus value, responding in the experimental condition was below the estimated baseline after removing carryover. A difference score of "0" means that after carry-over effects were considered, the performance of the experimental condition equaled the expected performance of the baseline condition. Positive difference scores reveal that responding in the experimental condition exceeded responding in the estimated baseline condition, after deleting the effects of carryover.

7. Team A was tested four times in the within-vehicle communications condition, producing difference scores of - 4.33, 0.17, -1.83, and -5.33.

8. Steps 1 through 7 were repeated for Teams B, C, and D, using words formed as the dependent variable. The four teams were tested a total of 14 times with the Sentence Construction Task in the within-vehicle communication condition. The average difference scores were less than what would have been expected during baseline, once carry-over effects were removed.

Mean Difference Score = Difference Score 1 +. . . . . .+ Difference Score N / (Number of Scores)
= (- 4.33)+(0.17)+ . . . . + (-1.83)+(-5.33) /14
= - 4.76

9. To allow comparison between dependent measures, all mean difference scores were computed as a percentage of the baseline mean. This metric will be called the mean deviation from baseline percentage (MDBP). The mean number of words formed for the four teams during baseline was 25.76. When crews communicated within the C^2V, their performances averaged 18% below baseline after removing carry-over effects.

MDBP = (Mean Difference Score / Baseline Performance) *100
= (-4.76 / 25.76) * 100
= -18.48

This correction for carryover assumes a linear relationship between the trials and group performance measures. Some investigators may object to this assumption because trial-performance functions are more likely to be negatively accelerated or negatively decelerated than linear (e.g., Mazur, 1994). Linearity was assumed in this control procedure in deference to simplicity. Any deviations from more complex functions that more precisely describe the trial-performance relationship should have a small effect on test results.

A more significant problem is that only three data points were used to compute the slope or best fitting line for each team. With only three data points, a single deviant score would have a pronounced effect on the slope. Fortunately, the problem caused by a lack of data points is attenuated because estimation errors should be randomly distributed. The probability of overestimating the slope in a positive direction should equal the probability of overestimating the slope in a negative direction. When slopes are computed for the four teams, slope estimation errors will tend to cancel.

The preceding plan for handling carry-over effects does not achieve the degree of control provided by complete counterbalancing or including trials as an independent variable. These control measures, which were devised for full experiments, cannot be applied to the LUT III group performance findings. Two options are available for analyzing the LUT III data. A less-than-ideal control procedure, such as the one recommended here, can be applied or the carry-over confound can be overlooked.

Unless the confound is treated, carryover could obscure the influence of the independent variables, leading investigators to conclude erroneously that experimental conditions had no differential effects on group performance. Also, performance differences attributable to carryover could incorrectly be ascribed to the experimental conditions. Therefore, the preceding control procedure will be applied to all LUT III group performance data.

Experimenter Effects

One person collected all the baseline data and other people supervised the data collection of crews housed in the vehicle. Would the data have been different if the individual who obtained baseline and the people who conducted in-vehicle testing switched roles? Ideally, investigators should have been shifted between the baseline and in-vehicle conditions.

Experimenter effects are well documented (e.g., Friedman, 1967; Rosenthal & Fode, 1963) in the behavioral sciences literature. Since all the data have been gathered, it is impossible to determine if the LUT III test administrators differentially affected crew performance. Often, when the experimenter does influence responding, his or her influence is a variable of minor importance (e.g., Barber, 1976). Hopefully, that is the case for the LUT III group performance tests. At this point, the only option is to proceed with the analysis under the assumption that test administrators had equivalent effects on crew performance.

Type of Communication-Competition Confound

In the intravehicular arrangement, four teammates were in the same C^2V. When tested in the intervehicular condition, each C^2V contained two participants from two different teams. As the evaluators intended, crew members housed in the same C^2V appeared to have an easier and friendlier communication environment. Unfortunately, the intervehicular condition contains a confound that could powerfully affect group performance.

Putting members from different teams in the same C^2V is a potentially competitive cue. No soldier needs to review the research literature (e.g., Beck & Pierce, 1996; Sherif, 1966) to appreciate the effects of competition on performance. Performance differences caused by the intravehicular versus intervehicular conditions could be attributed to either variations in the type of communication (between or within vehicle), co-action, rivalry, or a combination of variables. The effects of competition cannot be separated from the effects of intravehicular versus intervehicular communication.

Imposing Additivity on a Nonadditive World

This investigation examined the effects of vehicle movement, intravehicular versus intervehicular communication and terrain on group performance. Counting baseline, seven levels of the independent variables were manipulated. The inclusion of so many variables within the LUT III research design does not allow potentially important interactions to be examined.

For example, crews moving in a C^2V on Course A and communicating between vehicles (Moving-Between Vehicles-Course A) is one cell of the design. Each team received the Social Judgment Task only once in this combination of conditions. Any conclusions drawn from only one datum per team must be highly tentative.

Given the dearth of data, the only alternative is to collapse across conditions. For instance, performance on the 3-mile course and performance on Course A will be compared without taking the type of communication (intervehicle or intravehicle) into consideration. Summarizing across conditions is only appropriate when the effects of the independent variables are orthogonal or uncorrelated (Cook & Campbell, 1979). If effects are not orthogonal, assuming additivity loses information and distorts the relationships among independent variables.

A strong argument can be made that additivity of effects should not be assumed in this evaluation. To do so ignores a fundamental lesson of behavioral science. Social life is largely the product of interactions, many of them disordinal (e.g., Baron, Kerr, & Miller, 1992; Beck & Pierce, 1995). Apologies made, with so few observations per condition, the best choice is to assume additivity with reservations.

Tasks

For a detailed description of each task and possible dependent measures, see McGlynn et al. (in press).

Sentence Construction

The Sentence Construction Task was similar to one used by Crown and Rosse (1995). Each crew member received a different set of 27 letters from which he built words of three or more letters. For the first 5 minutes of the trial, soldiers worked without interacting with their teammates. Following the initial phase of the session, teams were allowed 25 minutes to form words into valid English sentences. Communication was permitted during this time. Instructions stipulated that a sentence must contain at least one word from each crew member. To facilitate sentence construction, crews were encouraged to trade letters to form words.

The number of sentences completed was the primary dependent variable. The number of words formed, letters used in making words, and letters traded with teammates were dependent measures of secondary importance. Highly cooperative teams should trade more letters, construct more words, and complete more sentences than less cooperative teams.

An examination of the data showed that teams frequently combined words into phrases that did not approximate sentences. Actual English sentences were the exception. Crews redesigned the task and in doing so, eliminated the main dependent variable. Without the central dependent measure, the analysis of the Sentence Construction Task was reduced to an examination of the number of words produced, letters used, and letters exchanged between teammates (see Figure 2 and Table 3).

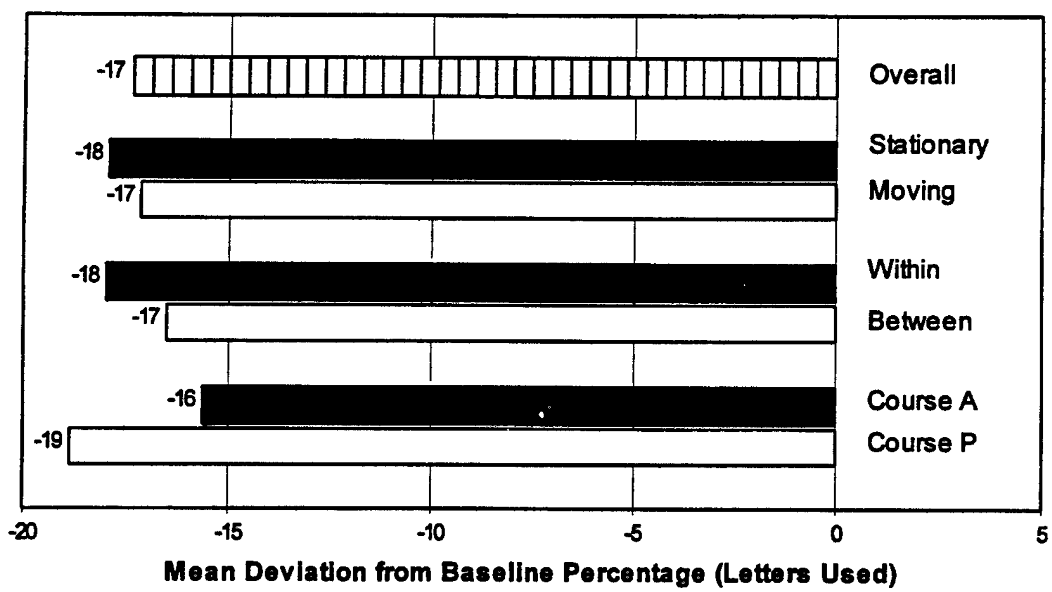

Figure 2. Letters used in the sentence construction task.

Table 3

Words Formed, Letters Used, and Letters Exchanged in the Sentence Construction Task

Independent variables	n	Words formed Total	MDBP[a]	Letters used Total	MDBP[a]	Letters exchanged Total	MDBP[a]
Baseline	12	6.44		23.06		3.71	
C^2V conditions	22	5.26	-18.35	19.03	-17.47	0.99	-73.20
Stationary	9	4.90	-24.03	18.92	-17.95	1.42	-61.59
Moving	13	5.28	-18.03	19.10	-17.14	0.70	-81.24
Between	8	4.88	-24.24	19.24	-16.54	3.40	-8.34
Within	14	5.26	-18.35	18.91	-18.00	-0.38	-110.26
Course A	7	5.34	-17.20	19.45	-15.66	1.47	-60.33
Course P	6	5.22	-19.01	18.71	-18.86	-0.21	-105.63

Note. MDBP=mean deviation from baseline percentage.
[a]Negative values indicate performances that are inferior to baseline.

Crews housed in C^2Vs scored below baseline in all experimental conditions. Impairments in the number of words produced (M = -18%) and letters used (M = -17%) were moderate in magnitude. Teams in C^2Vs rarely traded letters; the decline in performance averaged 73%. It is unlikely that the poor performance reflected by the letter exchange variable is solely attributable to the C^2V environment. Even in a difficult testing situation, motivated crews should have been more successful in trading letters. Perhaps the LUT III teams swapped so few letters because they were confused by the instructions or were disinterested in the task.

The number of words formed and letters used is essentially an individual measure with a collective component. Teams that actively exchange letters should increase the number of words they produce. The decline below baseline in words formed and letters used can largely be attributed to a failure of teammates to exchange letters when housed in the C^2V.

Social Judgment

Participants performing the Social Judgment Task (Beal, Gillis, & Stewart, 1978) are required to learn the relationship of predictor to criterion variables over a series of 15 problems. Each problem contains the same predictor and criterion variables. For example, the teams may use household income, the age of the car, and the education of the parents to estimate the number of miles that a family travels on vacation. For each problem, the crew members assess the importance of the predictors and estimate the criterion. After making their responses, the participants are told the actual criterion so that they may modify their responses to future problems.

Crew members worked alone and did not communicate during the first ten problems. For Problems 1 through 10, the investigators assigned the best predictor a statistical weighting of 70, the second best predictor a weighting of 50, and the least adequate predictor a weighting of 30. By the end of the tenth problem, team members were expected to learn that one variable (e.g., age of the car) is the most accurate predictor of the criterion (miles traveled). Teams were unaware that the predictor-criterion relationships were different for each crew member. For example, age of the car was the best predictor for one member of the team, and household income was the best predictor for another teammate.

The procedure was altered for the last five problems. Teams were instructed to discuss the importance of the predictors and to make a single estimate of the criterion. Also, for Problems 11 through 15, the relationship of the variables was changed so that the predictors were equally weighted.

The Social Judgment Task assumes that after the tenth problem, each crew member has a different opinion of what variable is the most useful predictor of the criterion. These opposing viewpoints should become apparent during the group discussion of Problems 11 through 15. To continue our example, one crew member should argue for stressing the age of the car and a teammate should emphasize household income in estimating the criterion. At this juncture, teams could either reconcile their differences or ignore the opinions of some teammates in making group decisions.

The assignment of weights to predictors and the estimation of the criterion reveal the extent of compromise in the final five problems. For instance, if all members' views were given equal consideration, teams should assign the same weights to all predictors in a given trial (e.g., car age = 33; household income = 33; parental education = 33; standard deviation (SD) = 0). Conversely, if the opinion of a single crew member predominates, there will be great variability in the weightings (e.g., car age = 75; household income = 15; parental education = 10; SD = 36.17).

This investigation used the SD of the weights as an index of variability. For the last five problems, the SD was computed, using the weights that the teams assigned to predictors as the data points. The mean SD was then calculated for Trials 11 through 15.

When housed in the C^2V, crews produced SDs that averaged 81% greater than baseline (see Figure 3 and Table 4). Standard deviations were particularly large when crews communicated within the vehicle or moved on the paved course. High SDs reveal unequal weights of predictors, implying that few viewpoints influenced the teams' decisions when crews were housed in the vehicle.

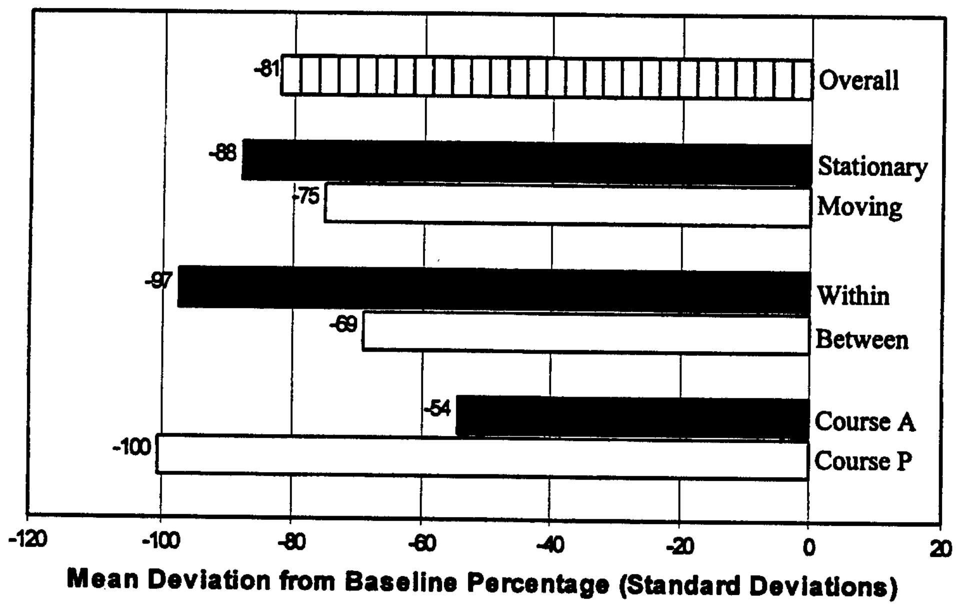

Figure 3. Standard deviations of teammates' weighting on the social judgment task.

Variability in the weights of predictors influences the accuracy of criterion estimates in the Social Judgment Task. Because the predictors were equally weighted during the final five trials, teams with low SDs in predictor weighting should give better estimates of the criterion than teams with higher SDs. In other words, teams that incorporate the views of all members in their criterion estimates should outperform teams that rely on only the opinions of one or two teammates. To test this hypothesis, the percentage of estimation error was calculated in Trials 11 to 15 using the following equation.

$$\text{Percentage Estimation Error}_{(\text{Trial } n)} = (\text{Absolute Value (Answer - Estimate) / Answer}) * 100$$

Table 4

Standard Deviations and Estimation Errors for the Social Judgment Task

Independent variables	n	SD Team	MDBP[a]	n	Estimation error Percentage	MDBP[a]
Baseline	10	17.45		8	16.02	
C^2V conditions	21	31.63	-81.25	17	23.01	-43.64
Stationary	10	32.79	-87.86	8	20.37	-27.13
Moving	11	30.59	-75.25	9	25.37	-58.33
Between	12	29.52	-69.14	9	19.84	-23.85
Within	9	34.45	-97.41	8	26.58	-65.92
Course A	6	26.93	-54.32	6	20.54	-28.21
Course P	5	34.97	-100.36	3	35.02	-118.56

Note. MDBP = mean deviation from baseline percentage.
[a]Negative values indicate performances that are inferior to baseline.

Overall, the criterion estimates of teams working in the C^2V were 44% less accurate than baseline (see Table 4 and Figure 4). Estimation errors were especially large when teams moved on the paved course and communicated within the same vehicle. Skepticism regarding the generality of this finding is warranted until the results can be replicated with a larger sample.

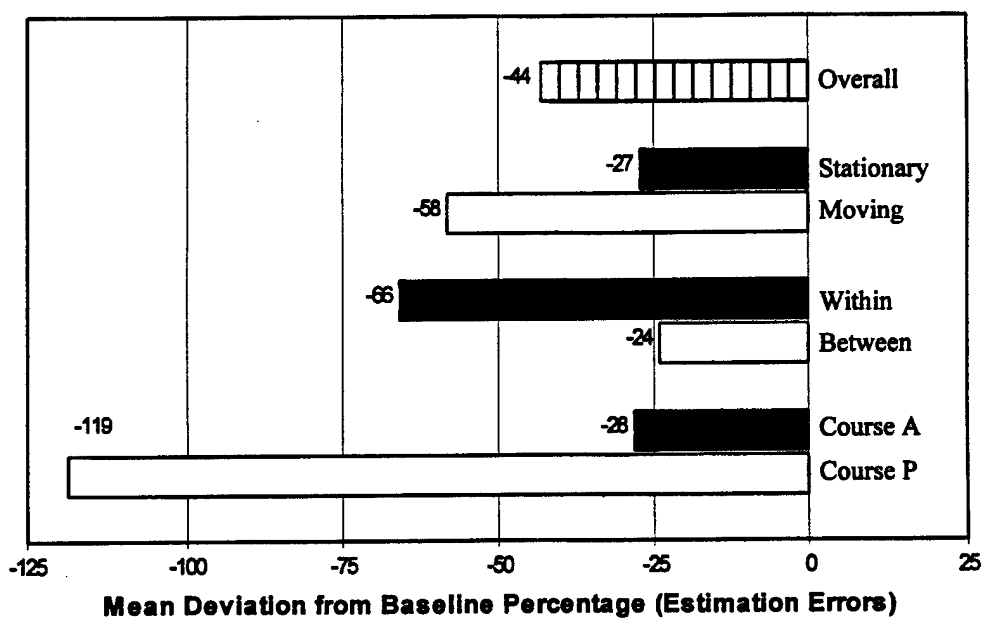

Figure 4. Estimation errors in the social judgment task.

In most respects, the Social Judgment data from the LUT III are consistent with the premise that placing crews in the C^2V reduces openness to the viewpoints of teammates. Restriction of the number of persons affecting decisions had an adverse impact, decreasing the accuracy of the teams' estimations of the criteria. Although the findings are in accord with this account, this explanation should be taken with caution. For the Social Judgment Task to work effectively, individuals must first have definite opinions about the relationships of the predictors to the criterion. Then, the team must forge a single decision from divergent opinions. An examination of the data from Problems 1 to 10 showed that many crew members failed to distinguish the associations of the predictor to the criterion variables. Some crew members began the last five problems with an opinion about the relative importance of predictors, but others were confused.

For some participants, the decisions made by the LUT III teams during the final five problems did not involve compromise because these individuals held no opinion. If the Social Judgment Task is to be used with similar participants in future evaluations, the weights must be made easier to discriminate. Instead of 30, 50, and 70, spread the weights to 15, 50, and 85. Furthermore, larger differences should be made in the criteria to simplify the social judgment problems. This suggestion is made with the wisdom of hindsight. *A priori* setting an effective difficulty level for a learning task is a very hazardous judgment.

Scrabble No. 2

Scrabble No. 2 was an adoption of the well-known parlor game and similar to a task used by McGlynn et al. (in press). At the beginning of a trial, each team member was given 40 letters, a list of letter point values, and a matrix with a seven-letter word in the center. Participants formed words from the letters and placed them on the matrix, following the usual Scrabble rules. After composing, the player communicated the word and its location on the matrix to his teammates. Teams were encouraged to trade letters to form more words. Whenever a crew member received a letter, he was required to give a letter from his set to his teammate. Four 4-minute tests were conducted during each session. The dependent variables were number of words formed, letters used, and points obtained averaged over the four tests.

Performance of the Scrabble No. 2 Task was below baseline for all conditions in which crews worked in the C^2V (see Figure 5 and Table 5). When tested in the vehicle, crews composed 20% fewer words, used 6% fewer letters, and obtained 12% fewer points. Performance was particularly low if trials were conducted on Course A.

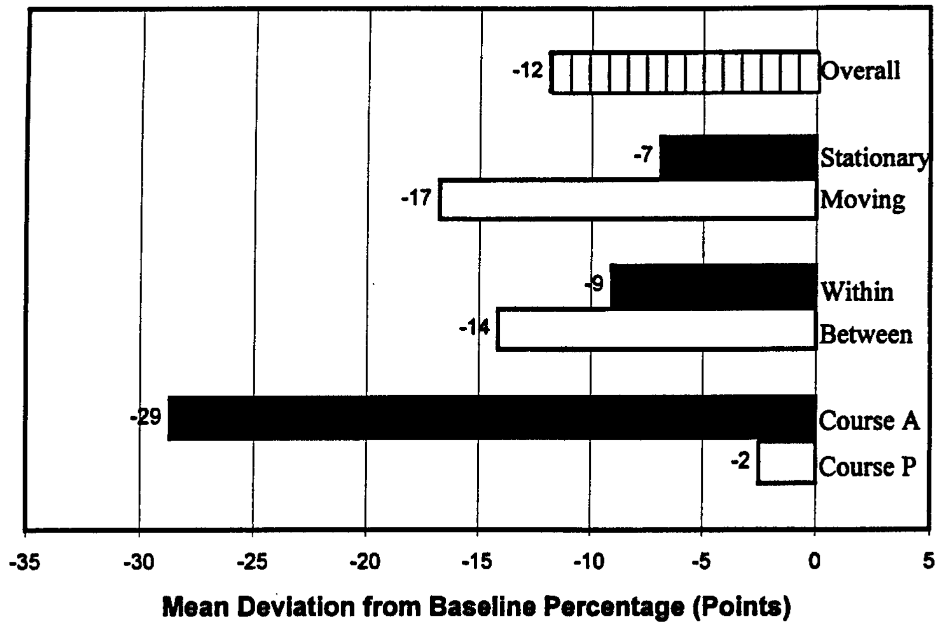

Figure 5. Points gained on the Scrabble No. 2 task.

Table 5

Words Formed, Letters Used, and Points Gained on the Scrabble No. 2 Task

Independent variables	n	Words formed Total	Words formed MDBP[a]	Letters used Total	Letters used MDBP[ab]	Points gained Total	Points gained MDBP[a]
Baseline	11	8.03		34.77		83.35	
C^2V conditions	22	6.46	-19.60	32.69	-5.99	73.48	-11.84
Stationary	11	6.57	-18.18	33.46	-3.78	77.60	-6.90
Moving	11	6.34	-21.02	31.92	-21.02	69.36	-16.78
Between	12	6.43	-19.90	29.42	-15.39	71.56	-14.14
Within	10	6.48	-19.24	36.61	5.29	75.78	-9.08
Course A	6	5.69	-29.11	29.58	-14.94	59.44	-28.69
Course P	5	7.12	-11.31	34.74	-0.10	81.27	-2.50

Note. MDBP = mean deviation from baseline percentage.

[a]Negative values indicate performances that are inferior to baseline.

[b]Positive values show performances that are superior to baseline.

Quiz Task

The Quiz Task was modeled after a task used by Littlepage and Silbiger (1992). The LUT III teams were given ample time to complete 20 recall items. Questions were taken from a variety of topics (e.g., sports, history, entertainment) to increase the likelihood that each soldier would know some answers. After responding to a question, the teams rated their confidence in the correctness of the chosen answer. A 100-point confidence scale was employed with high scores showing the greatest confidence. Teams were permitted to discuss each question before settling on an answer and confidence rating. The mean number of correct responses and the mean confidence score were the primary dependent variables.

One team's confidence data were unusual; their ratings were almost as high for the items they missed as for the items that they correctly answered. In this investigator's opinion, it is improbable that this team was completely unaware of what information they knew. More likely, they did not understand the instructions, were inattentive in completing the ratings, or were reluctant to admit that they were unsure of some answers. Therefore, this team's confidence data were considered invalid and were deleted from the analysis.

Performance in the C^2V approximated baselines on the Quiz Task (see Figure 6 and Table 6). The C^2V environment had a much smaller effect on the Quiz Task than on other tasks included in the LUT III. When the four tasks are considered together, the data reveal that working in a C^2V impairs some, but not all, assignments that teams perform.

Investigations should be conducted to determine the types of tasks that are especially likely to be imparied by housing crews in a C^2V. One straightforward hypothesis is that the C^2V has a more destructive impact on the collective than the individual components of team performance. If this proposition is correct, then tasks that put a premium on group processes should be most adversely affected by the C^2V environment. A post hoc examination of the LUT III data provides some support for this proposition. Only a minimal degree of interaction is necessary for a team to do well on the Quiz. In comparison, the Sentence Construction, Social Judgment, and Scrabble 2 Tasks appear to require more complex forms of social interaction.

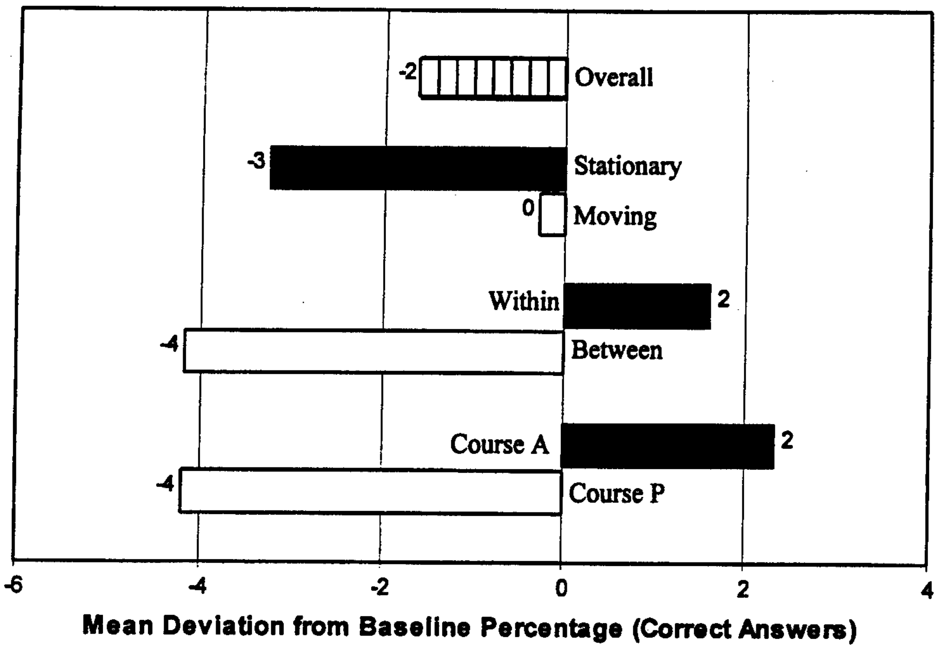

Figure 6. Correct answers to the quiz task.

Table 6

Confidence Ratings in and Correct Answers to the Quiz Task

Independent variables	n	Confidence ratings Mean	Confidence ratings MDBP[a]	Correct answers Total	Correct answers MDBP[bc]
Baseline	12	75.22		12.92	
C^2V conditions	18	78.00	3.70	12.71	-1.60
Stationary	8	77.79	3.41	12.50	-3.26
Moving	10	78.18	3.93	12.88	-0.28
Between	10	75.84	0.82	12.38	-4.17
Within	8	80.71	7.30	13.12	1.61
Course A	6	79.04	5.08	13.22	2.33
Course P	4	76.89	2.21	12.38	-4.19

Note. MDBP = mean deviation from baseline percentage.
[a]Positive values reveal confidence greater than baseline.
[b]Negative values indicate performances that are inferior to baseline.
[c]Positive values show performances that are superior to baseline.

SUMMARY

Constructing a Summary Metric

Besides examining team performance at the level of functions, assessing the overall effect of variables is often valuable. People frequently need to know the general or averaged effects of the experimental manipulation on performance. They are seeking a more global answer than any single task can provide.

The most significant benefit in computing a summary or overall performance index is that it provides a method for examining important interactions between independent variables. Interpreting interactions on any single LUT III task would be clearly inappropriate because so few observations per cell were recorded. Assessments based on one or two observations per cell would probably lead to some unusual and misleading conclusions. However, because an overall measure is calculated from performances in many different tests, the total number of observations per cell is increased. A summary measure of performance offers the potential for examining important dependencies, such as the interaction of type of communication (intravehicle versus intervehicle) and movement (stationary versus moving).

Whenever the results of a series of molecular tasks are to be combined to form a molar metric, the question of how each index should be weighted must be considered (e.g., Tabachnick & Fidell, 1989). Empirical studies have not yet revealed what group tasks best discriminate successful from unsuccessful C^2V crews. For example, no investigation has compared the predictive utilities of the Social Judgment and Sentence Construction Tasks. Given the current state of knowledge, the most reasonable approach is to equally weight each task.

Most of these tasks yield multiple dependent indices. Not all the data were valid. For instance, the number of sentences formed in the Sentence Construction Task was an invalid index. From the valid measures, these evaluators selected the dependent variable that they felt was the most important performance index for each task. These were Sentence Construction, letters used; Scrabble No. 2, points scored; Social Judgment, estimated error; and Quiz, accuracy.

Another problem in developing a summary measure is that the tasks yield very different indices. How can total points scored during Scrabble No. 2 be added to accuracy data from the Quiz Task? A common stratagem is to calculate a standard score for each test before summing. The small LUT III sample does not permit the use of standard scores. The next best alternative is to compute the summary performance measure from the MDBP scores.

Computing a summary performance index is exacerbated by the number of independent variables included in LUT III. Besides a baseline, the design held six cells: stationary--between vehicle communication, stationary--within vehicle communication, moving--between vehicle communication-Course A, moving--between vehicle communication-3-mile course, moving--within vehicle communication-Course A, and moving--within vehicle communication-3-mile course.

Crew members were scheduled for nine trials for most tasks; three of the trials were baselines. Ideally, each team could be tested once in the six remaining conditions. Because of equipment malfunctions and data collection difficulties, some cells contain no observations. Probably, the best way to handle this obstacle is to collapse across the least important independent variable, terrain. This yields a design with a baseline and four experimental cells: stationary--between vehicle communication, stationary--within vehicle communication, moving--between vehicle communication, and moving--within vehicle communication. With few exceptions, teams were tested at least once in each of these conditions.

The summary score for a particular condition was the average MDBP score. For example, the four teams conducted a total of 19 trials in which the C^2V was stationary and communication was within vehicle. The summary score was obtained by (a) multiplying the number of stationary--within vehicle trials for each task by the corresponding MDBP score, (b) summing across the four tasks, and (c) dividing by the total number of stationary--within vehicle trials.

$$\begin{aligned}\text{Stationary-Within} &= \{(\text{MDBP}_{\text{Sen Con}} * \text{Trials}_{\text{Sen Con}}) + \ldots\ldots + (\text{MDBP}_{\text{Quiz}} * \text{Trials}_{\text{Quiz}}) \,/\, \text{Total Trials}\} \\ &= \{(-23.41 * 6) + \ldots\ldots + (-4.31 * 5) \,/\, 19\} \\ &= -18.11\end{aligned}$$

If LUT III were a full scale evaluation, summary scores would not be computed in this manner. The recommended procedures are an effort to construct an overall performance metric that can be applied to small samples. Still, the summary method is a far better approach than for the evaluator to weight tasks and to form a subjective conclusion about the overall performance.

Summary Measure Results

As Figure 7 and Table 7 show, the performances of crews housed in the C^2V averaged 18% below the performances of teams operating in baseline conditions. The impairment of group functioning in the C^2V cannot be solely attributed to movement. Even when the vehicle was stationary, performance scores averaged 13% less than the baseline. Examination of the means suggests a slight interaction, in which the effects of communication type (intervehicle, intravehicle) were slightly greater if the vehicle was stationary.

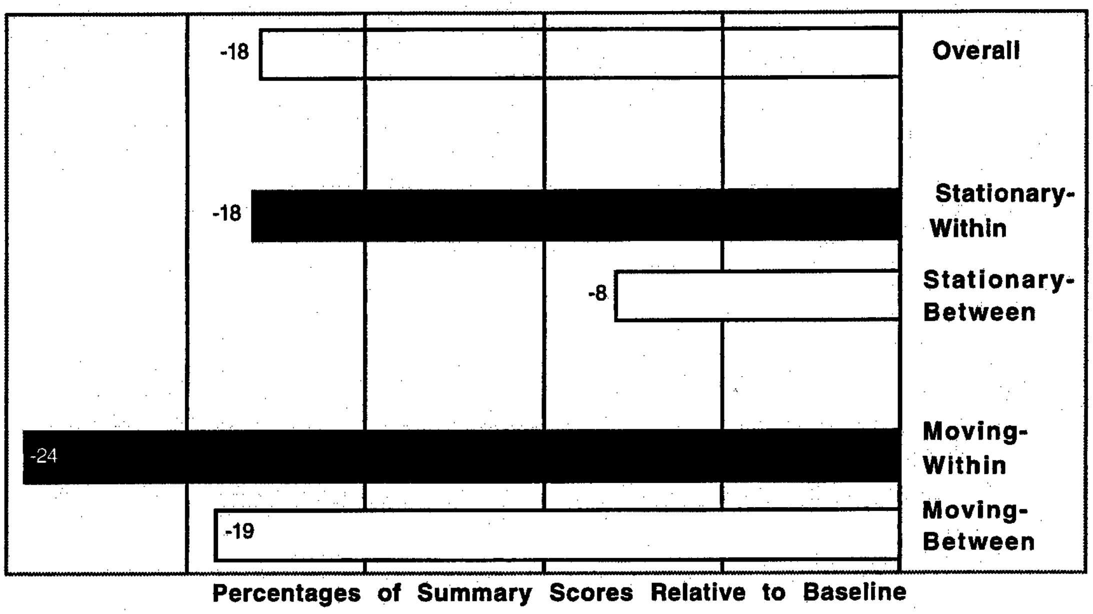

Figure 7. The effects of movement and intravehicular versus intervehicular communication.

Table 7

Summary Scores of the C^2V Conditions

	Stationary	Moving	Totals
Within vehicle	-18.11 (19)	-24.43 (21)	-21.43 (40)
Between vehicle	-8.02 (17)	-19.18 (22)	-14.32 (39)
Totals	-13.34 (36)	-21.74 (43)	-17.91 (79)

Note. Summary scores are the average of the mean deviation from baseline percentages. Negative summary scores indicate performances that are inferior to baseline. The number of trials in each condition is in parentheses.

The discovery that stationary C^2Vs caused a decline in team performance is an important finding that merits further inquiry. The C^2V environment contains many potentially powerful debilitators that could affect performance when the vehicle is in a stationary posture. For instance, performance in stationary C^2Vs may have been below the baseline because (a) audio provided by the intercom and SINCGARS was inferior to voice, (b) the seating of the crew in the vehicle restricted visual cues, (c) the LUT III crews were inexperienced with C^2Vs, or (d) noise, heat, and other distracters in the vehicle disrupted communication. A series of investigations will ultimately be needed to discover if these or other factors impair the ability of C^2V crews to integrate their activities.

Averaged over all tasks, movement augmented the detrimental impact of the C^2V environment on performance ($\underline{M}_{\text{Stationary}}$ = -14%; $\underline{M}_{\text{Moving}}$ = -22%). For safety reasons, the C^2V was restricted to a top speed of 20 mph. When fielded, teams will sometimes need to conduct C2 when the vehicle is exceeding 40 mph. The effect of high speeds on group performance is a topic for future research.

The most surprising outcome was that performance was better when crew members communicated between ($\underline{M}$ = -14%), rather than within vehicles, ($\underline{M}$ = -21%). Putting teammates in adjacent C^2Vs would presumably create a barrier to communication, potentially disrupting performance. Why should crews housed in separate vehicles perform better than crews working in the same vehicle? The most probable explanation is that the between-vehicle manipulation contained a serious confound. Two teammates worked alongside two members of another team in the intervehicular condition. The presence of persons from other teams was probably a stimulus for competition. In comparison with the intravehicular arrangement, the intervehicular manipulation obstructed communication but heightened competition (co-action and rivalry).

During the LUT III, the beneficial effects of increased competition outweighed the negative impact on communication. Teams performed better in the between-vehicle than the within-vehicle condition. Although this account is consistent with the results, such post hoc explanations are never fully satisfying. Other plausible interpretations could be offered. A better understanding of the effects of intravehicular versus intervehicular communication will not be obtained until evaluations are designed without confounds in critical independent variables.

What do the LUT III group task data suggest about the performance of C^2V crews in the field? To make this extrapolation, the testing situation must be compared with the actual conditions C^2V crews will encounter. With few exceptions, LUT III teams were not pressured to process or trade information rapidly. If the crew wanted, messages could be repeated to ensure comprehension. Even the slowest LUT III teams completed most tasks in the given time. Successful integration of crew members' activities resulted in better group performance, but a high degree of efficiency was not required to do the tasks well.

The LUT III data best generalize to situations in which the team is given ample time to process information and is either stationary or moving at a moderate speed in the vehicle. Military teams often work in conditions such as these. However, the LUT III findings reveal little about how C^2V crews will respond to severe time dictates or how they will conduct C2 when the vehicle is moving at top speed. Additional research is required to determine if C^2V crews can successfully coordinate during fast paced activities on the battlefield or in other challenging environments.

The principal findings of the LUT III group performance tests were

• The performance of crews housed in C^2Vs was inferior to that of teams working in benign baseline conditions. Even when the vehicle was stationary, teams performed below baseline.

• Movement increased the detrimental effect of the C^2V environment on team performance. Crew performance scores averaged 14% below the baseline when the vehicle was stationary, compared to 21% below the baseline when the vehicle was moving.

• In three of four tasks, performance was better on Course A than on the paved 3-mile course. Superiority of Course A was only pronounced on the Social Judgment Task. The small sample size suggests caution in making any conclusions regarding the effects of terrain on group performance.

• The effects of working in the C^2V on performance depended on the task. The C^2V environment had a significant detrimental impact on performance of the Sentence Construction, Social Judgment, and Scrabble 2 Tasks. However, performance in the C^2V approximated baseline on the Quiz Task. One interpretation of these results is that the C^2V environment most adversely affects the performance of tasks that stress the importance of crews integrating their activities.

• The results of the LUT III group performance tests should best generalize to situations in which crews do not need to rapidly transmit or process information. The findings are also most applicable to circumstances in which the vehicle is moving at slow to moderate speeds.

Investigative Issues

If the C^2V is to become a prominent part of the 21st century Army's arsenal, then it should be developed so as to maximize group task performance as an analog to command staff performance. The concluding section of this evaluation examines the requirements of future C^2V team performance tests:

- Measuring team performance in more technologically advanced C^2Vs.
- Testing crews in conditions more challenging than those used in the LUT III.
- Providing sufficient resources to assess for interactions between the variables that control collective behavior.
- Empirically establishing the relationship between a set of team tasks and group functions.

Technological Innovation and Group Performance

C^2V crews must transmit and process more information than current command posts do. This increase in workload must be accomplished with fewer personnel. Advanced electronic technologies are expected to improve efficiency, enabling C^2V crews to handle high rates of information input. Fielded C^2Vs will be equipped with intelligent software for searching, sorting, prioritizing, and transmitting. Enhanced audio and video communication instruments, flat screen monitors, and electronic battle maps are other devices that will presumably facilitate C2.

The prototypes used in the LUT III lacked most of the electronics that will someday be the heart of the C^2V's communication system. The ATCCS equipment was unfortunately not available for the LUT III. Each major technological innovation will solve some problems and create others. As the technology changes, so will the optimum interaction patterns between humans and between humans and machines. Technological innovation will be a driving force, requiring many group performance studies.

Challenge and Nonadditivity

One of the most important lessons of social psychology is that the variables that determine group performance often combine nonadditively. In other words, the combined effects of independent variables on performance could not be predicted from studying any independent variable in isolation. Unfortunately, testing for interactions requires larger samples than testing for main effects. Small sample evaluations, such as the LUT III, can yield partial or misleading pictures of the effects of variables because they do not allow investigators to measure nonadditive relations. Until resources are available for larger studies, the Army will have a very limited knowledge of the variables that determine how C^2V crew members combine skills and synchronize their activities.

Testing C^2V crews under various levels of stress will reveal a series of important interactions. For example, if vehicle speed affects environmental stress, clear predictions can be made from the behavioral sciences literature (e.g., Hull, 1943). Increases in vehicle speed will produce pronounced impairments in the performance of cognitively complex or novel tasks. The deleterious effects of high speed will be much smaller if the task is simple or well practiced. If the primary effect of vehicle speed is upon the individual's arousal level, crews may perform simple C2 tasks better at fast speeds than at slow speeds.

Terrain may also be interactive, a trivial variable at slow speeds but a more important variable at higher speeds. To extrapolate from the social psychology literature (e.g., Zajonc, 1980), terrain effects will also depend on the type of task. Terrain may have little influence on tasks that are neither physically nor cognitively demanding but will have a significant impact on more difficult assignments.

Comparisons of intravehicular versus intervehicular communications must also consider dependencies among the independent variables. In well-controlled experiments, the effects of between- versus within-vehicle communication are likely to increase as a function of the demands made upon the crew. Communicating between vehicles may have little effect in low demand conditions but may have serious deleterious effects if tasks are complex, vehicle speeds are high, or time is restricted.

The combination of social psychological variables will have a powerful impact on the effectiveness of C^2V teams. Communication networks, information filtering, leader-subordinate relations, diffusion of responsibility, free riding, and equity are some group processes that will interact with the type of task to control group performance. For instance, if information is received at a slow or moderate rate, C^2V crews configured in a centralized network will probably outperform decentralized crews. However, if the rate of information flow increases, decentralized teams will be more effective than centralized networks (Beck & Pierce, 1995). Disordinal relationships between social and environmental variables are common place and their elucidation will be fundamental to the development of efficient C^2V teams.

Empirically Based Group Performance Battery

For many years, group process researchers have stressed the importance of empirically deriving a set of group performance functions (e.g., Hackman & Morris, 1975). Ideally, a battery of tasks should be identified that provide accurate measures of these functions.

McGlynn's work, based on Fleishman and Zaccaro's (1992) functions, is progress in the right direction.

The main shortcoming of the LUT III battery was that the tasks were logically, rather than empirically, related to group function. Given that the association of performance tests to group functions is probably highly complex, any logically derived set of tasks and functions should be suspect. A series of empirical investigations may reveal different factors than McGlynn proposed. Also, tasks will probably be sensitive to multiple functions, and this interactivity will need to be considered in any application of the test battery.

A methodology is proposed in hopes of stimulating investigators to develop a test battery that is empirically related to group functions. The procedure is an adoption from psychometric test and questionnaire construction procedures (e.g., Anastasi, 1988; Spector, 1992). The basic methodology is a well-worn psychometric path, but researchers will confront problems that are idiosyncratic to the development of a group performance test battery.

The central difference in the validation of a group test battery and most psychometric instruments is the unit assessed for inclusion. Most psychometric tests begin with a sample of items from which a subset of empirically derived questions is identified. Development of a group performance battery begins with a pool of tasks from which valid estimators of the constructs are chosen. The establishment of an empirically grounded group test battery should follow these steps.

1. The selection of group performance tests must be preceded by the identification of a set of hypothesized team functions. McGlynn's functions are an example of this first step in test battery development. A research team now needs to reexamine McGlynn's modification of Fleishman and Zaccaro's functions, taking the LUT III data into consideration. The team may decide to continue with McGlynn's taxonomy or modify the list of functions.

2. Several tasks should be chosen for each hypothesized function. Multiple tasks are needed because the loadings of particular tasks on functions cannot be predicted with certainty. Many studies (e.g., Ingham, Levinger, Graves, & Pickham, 1974; Kerr, 1989) have shown that the number of participants affects performance, so group size must be taken into consideration in constructing the test battery. Unless one is willing to conduct a separate study for each group size, the number of participants must be held constant across tasks. To increase generality, it is recommended that all tasks be designed for a moderate sized group. Many social

phenomena can be shown with four-person teams, and four is a manageable group for most experimental settings.

3. Participants must be adults and of at least of average intelligence. With these stipulations, external validity can be enhanced by building heterogeneity into the testing sample. If a useful group performance battery is to be developed, it is vital that eventually each participant be tested in every task.

4. A benign testing environment, similar to the baseline condition used in LUT III, must be established. Besides relating tasks to function, this investigation will provide baseline norms for each task.

5. Experienced applied social or organizational psychologists will be needed to design the specifics of the project. The primary investigator must also have a strong background in psychometrics. Persons not specifically trained in social or organizational psychology can be used in test delivery and data compilation. However, it is highly unlikely that minimal standards of scientific credibility will be achieved unless a social or organizational psychologist is at the helm.

6. The raw data will be the scores that teams receive on each test. In the analysis, each team performance measure will be treated similarly to an item on a questionnaire or ability test. No team performance measure will be assumed to be more significant or weighted more than any other team performance measure.

7. A factor analysis of the data will be conducted. This will yield a set of group functions and one or more tests that are measures of that function. No *a priori* rationale suggests that the factors or functions will be orthogonal. Therefore, a nonorthogonal factor analysis will first be performed. If the solution suggests a high degree of independence between factors, a varimax or other nonorthogonal solution will be attempted.

The lack of a group test battery that empirically connects tasks to group functions is probably the greatest impediment to understanding the collective behavior of C^2V and other Army crews. Until such a battery is developed, knowing with certainty that a comprehensive assessment of team functioning has been conducted will be impossible. For too long, logic has been allowed to substitute for real data. Now is the time to initiate a series of investigations that will culminate in a test battery that is empirically tied to team performance functions.

REFERENCES

Anastasi, A. (1988). Psychological testing. (5th ed.). New York: Macmillan.

Barber, T.X. (1976). Pitfalls in human research: ten pivotal points. New York: Pergamon Press.

Baron, R.S., Kerr, N.L., & Miller, N. (1992). Group process, group decision, group action. Pacific Grove, CA: Brooks/Cole.

Beal, D., Gillis, J.S., & Stewart, T. (1978). The lens model: Computational procedures and applications. Perceptual and Motor Skills, 46, 3-28.

Beck, H.P., & Pierce, L.G. (1995). The effects of group processes and information rate on the performance of C2V crews: An experimental proposal (unpublished manuscript). Fort Sill, OK: U.S. Army Research Laboratory.

Beck, H.P., & Pierce, L.G. (1996). The impact of selected group processes on the coordination and motivation of Army teams (ARL-CR-292). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Christensen, L.B. (1980). Experimental methodology. Boston: Allyn & Bacon.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton Mifflin.

Crown, D.F., & Rosse, J.G. (1995). Yours, mine, and ours: Facilitating group productivity through the integration of individual and group goals. Organizational Behavior and Human Decision Processes, 64, 138-150.

Edwards, A.L. (1968). Experimental design in psychological research. New York: Holt, Rinehart, & Winston.

Fleishman, E.A., & Zaccaro, S.J. (1992). Toward a taxonomy of team performance functions. In R.W. Swezey & E. Salas (Eds.), Teams: Their training and performance (pp. 31-56). Norwood, NJ: Ablex.

Friedman, N. (1967). The social nature of psychological research. New York: Basic Books.

Hackman, J.R., & Morris, C.G. (1975). Group tasks, group interaction process, and group performance effectiveness. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 8). New York: Academic Press.

Himmelfarb, S. (1975). What do you do when the control group doesnít fit into the factorial design? Psychological Bulletin, 82, 363-368.

Hull, C.L. (1943). Principles of behavior. New York: Appleton-Century-Crofts.

Ingham, A.G., Levinger, G., Graves, J., & Pickham, V. (1974). The Ringelmann effect: Studies of group size and group performance. Journal of Experimental Social Psychology, 10, 371-384.

Isenberg, D.J. (1981). Some effects of time-pressure on vertical structure and decision-making accuracy in small groups. Organizational Behavior and Human Performance, 27, 119-134.

Kerr, N.L. (1989). Illusions of efficacy: The effects of group size on perceived efficacy in social dilemmas. Journal of Experimental Social Psychology, 25, 287-313.

Littlepage, G.E., & Silbiger, H. (1992). Recognition of expertise in decision making groups: Effects of group size and participation patterns. Small Group Research, 23, 344-355.

Laughlin, P.R., VanderStoep, S.W., & Hollingshead, A.B. (1991). Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. Journal of Personality and Social Psychology, 61, 50-67.

Martin Marietta Energy Systems (1993). Future command and control vehicle: Functional assessment. Leavenworth, KS: Author.

Mazur, J.E. (1994). Learning and behavior. Englewood Cliffs, NJ: Prentice-Hall.

McGlynn, R.P., Sutton, J.L., Demski, R.M., Sprague, V.L., & Pierce, L.P. (in press). Development of a team performance task battery to evaluate performance of the command and control vehicle (C^2V) crew (Contract No. DAAL01-96-P-0875). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Rosenthal, R., & Fode, K.L. (1963). Three experiments in experimenter bias. Psychological Reports, 12, 491-511.

Sherif, M. (1966). In common predicament: Social psychology of intergroup conflict and cooperation. Boston: Houghton Mifflin.

Spector, P.E. (1992). Summated rating scale construction. Newbury Park, CA: Sage.

Stasser, G., Taylor, L.A., & Hanna, C. (1989). Information sampling in structured and unstructured discussions of three- and sex-person groups. Journal of Personality and Social Psychology, 57, 67-78.

Tabachnick, B.G., & Fidell, L.S. (1989). Using multivariate statistics. Cambridge, UK: Harper & Row.

Zajonc, R.B. (1980). Compresence. In P.B. Paulus (Ed.), Psychology of group influence (pp. 35-60). Hillsdale, NJ: Erlbaum.

NO. OF COPIES	ORGANIZATION
2	ADMINISTRATOR DEFENSE TECHNICAL INFO CENTER ATTN DTIC DDA 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CS AL TA RECORDS MANAGEMENT 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CI LL TECHNICAL LIBRARY 2800 POWDER MILL RD ADELPHI MD 207830-1197
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CS AL TP TECH PUBLISHING BRANCH 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTORATE FOR MANPRINT ATTN DAPE MR DEPUTY CHIEF OF STAFF PERSONNEL 300 ARMY PENTAGON WASHINGTON DC 20310-0300
1	DIRECTOR ARMY AUDIOLOGY & SPEECH CENTER WALTER REED ARMY MED CENTER WASHINGTON DC 20307-5001
1	OUSD(A)/DDDR&E(R&A)/E&LS PENTAGON ROOM 3D129 WASHINGTON DC 20301-3080
1	CODE 1142PS OFFICE OF NAVAL RESEARCH 800 N QUINCY STREET ARLINGTON VA 22217-5000
1	WALTER REED ARMY INST OF RSCH ATTN SGRD UWI C (COL REDMOND) WASHINGTON DC 20307-5100

NO. OF COPIES	ORGANIZATION
1	COMMANDER US ARMY RESEARCH INSTITUTE ATTN PERI ZT (DR E M JOHNSON) 5001 EISENHOWER AVENUE ALEXANDRIA VA 22333-5600
1	DEPUTY COMMANDING GENERAL ATTN EXS (Q) MARINE CORPS RD&A COMMAND QUANTICO VA 22134
1	HEADQUARTERS USATRADOC ATTN ATCD SP FORT MONROE VA 23651
1	COMMANDER US ARMY MATERIEL COMMAND ATTN AMCAM 5001 EISENHOWER AVENUE ALEXANDRIA VA 22333-0001
1	COMMANDER USA OPERATIONAL TEST & EVAL AGENCY ATTN CSTE TSM 4501 FORD AVE ALEXANDRIA VA 22302-1458
1	USA BIOMEDICAL R&D LABORATORY ATTN LIBRARY FORT DETRICK BUILDING 568 FREDERICK MD 21702-5010
1	COMMANDER USA AEROMEDICAL RESEARCH LAB ATTN LIBRARY FORT RUCKER AL 36362-5292
1	AAMRL/HE WRIGHT PATTERSON AFB OH 45433-6573
1	US ARMY TROOP SUPPORT CMD NATICK RD&E CENTER ATTN BEHAVIORAL SCI DIV SSD NATICK MA 01760-5020
1	US ARMY TROOP SUPPORT CMD NATICK RD&E CENTER ATTN TECH LIBRARY (STRNC MIL) NATICK MA 01760-5040

NO. OF COPIES	ORGANIZATION
1	MEDICAL LIBRARY BLDG 148 NAVAL SUBMARINE MEDICAL RSCH LAB BOX 900 SUBMARINE BASE NEW LONDON GROTON CT 06340
1	DR JON FALLESEN ARI FIELD UNIT PO BOX 3407 FORT LEAVENWORTH KS 66027-0347
1	ARI FIELD UNIT FORT KNOX BUILDING 2423 PERI IK FORT KNOX KY 40121-5620
1	COMMANDANT USA ARTILLERY & MISSILE SCHOOL ATTN USAAMS TECH LIBRARY FORT SILL OK 73503
1	COMMANDER WHITE SANDS MISSILE RANGE ATTN TECHNICAL LIBRARY WHITE SANDS MISSILE RANGE NM 88002
1	COMMANDER USA TANK-AUTOMOTIVE R&D CENTER ATTN AMSTA TSL (TECH LIBRARY) WARREN MI 48397-5000
1	PEO COMMUNICATIONS ATTN SFAE CM RE FT MONMOUTH NJ 07703-5000
1	COMMANDANT US ARMY ARMOR SCHOOL ATTN ATSB CDS (MR LIPSCOMB) FT KNOX KY 40121-5215
1	CECOM SP & TERRESTRIAL COM DIV ATTN AMSEL RD ST MC M H SOICHER FT MONMOUTH NJ 07703-5203
1	PRIN DPTY FOR TECH GY HDQ US ARMY MATL CMND ATTN AMCDCG T M FISETTE 5001 EISENHOWER AVE ALEXANDRIA VA 22333-0001
1	PRIN DPTY FOR ACQTN HDQ US ARMY MATL CMND ATTN AMCDCG A D ADAMS 5001 EISENHOWER AVE ALEXANDRIA VA 22333-0001

NO. OF COPIES	ORGANIZATION
1	DPTY CG FOR RDE HDQ US ARMY MATL CMND ATTN AMCRD BG BEAUCHAMP 5001 EISENHOWER AVE ALEXANDRIA VA 22333-0001
1	DPTY ASST SCY FOR RSRCH & TECH SARD-TT F MILTON RM 3E479 THE PENTAGON WASHINGTON DC 20310-0103
1	ODCSOPS D SCHMIDT WASHINGTON DC 20310-1001
1	OSD OUSD(A&T)/ODDDR&E(R) J LUPO THE PENTAGON WASHINGTON DC 20301-7100
1	ARL ELECTROMAG GROUP CAMPUS MAIL CODE F0250 A TUCKER UNIVERSITY OF TEXAS AUSTIN TX 78712
1	DUSD SPACE 1E765 J G MCNEFF 3900 DEFENSE PENTAGON WASHINGTON DC 20301-3900
1	USAASA MOAS-AI W PARRON 9325 GUNSTON RD STE N319 FT BELVOIR VA 22060-5582
1	CECOM PM GPS COL S YOUNG FT MONMOUTH NJ 07703
1	GPS JOINT PROG OFC DIR COL J CLAY 2435 VELA WAY STE 1613 LOS ANGELES AFB CA 90245-5500
1	ELECTRONIC SYSTEMS DIV DIR CECOM RDEC J NIEMELA FT MONMOUTH NJ 07703
3	DARPA L STOTTS J PENNELLA B KASPAR 3701 N FAIRFAX DR ARLINGTON VA 22203-1714

NO. OF COPIES	ORGANIZATION
1	SPECIAL ASST TO THE WING CDR 50SW/CCX CAPT P H BERNSTEIN 300 O'MALLEY AVE STE 20 FALCON AFB CO 80912-3020
1	USAF SMC/CED DMA/JPO M ISON 2435 VELA WAY STE 1613 LOS ANGELES AFB CA 90245-5500
1	ARL HRED AVNC FIELD ELEMENT ATTN AMSRL HR MJ (R ARMSTRONG) PO BOX 620716 BLDG 514 FT RUCKER AL 36362-0716
1	ARL HRED MICOM FIELD ELEMENT ATTN AMSRL HR MO (T COOK) BUILDING 5400 ROOM C242 REDSTONE ARSENAL AL 35898-7290
1	ARL HRED USAADASCH FLD ELEMENT ATTN AMSRL HR ME (K REYNOLDS) ATTN ATSA CD 5800 CARTER ROAD FORT BLISS TX 79916-3802
1	ARL HRED ARDEC FIELD ELEMENT ATTN AMSRL HR MG (R SPINE) BUILDING 333 PICATINNY ARSENAL NJ 07806-5000
1	ARL HRED ARMC FIELD ELEMENT ATTN AMSRL HR MH (J JOHNSON) BLDG 1109B 3RD FLOOR FT KNOX KY 40121-5215
1	ARL HRED CECOM FIELD ELEMENT ATTN AMSRL HR ML (J MARTIN) MYER CENTER RM 3C214 FT MONMOUTH NJ 07703-5630
1	ARL HRED FT BELVOIR FIELD ELEMENT ATTN AMSRL HR MK (P SCHOOL) 10115 GRIDLEY ROAD SUITE 114 FORT BELVOIR VA 22060-5846
1	ARL HRED FT HOOD FIELD ELEMENT ATTN AMSRL HR MV (E SMOOTZ) HQ TEXCOM BLDG 91012 RM 111 FT HOOD TX 76544-5065

NO. OF COPIES	ORGANIZATION
2	ARL HRED NATICK FIELD ELEMENT ATTN AMSRL HR MQ (M FLETCHER) ATTN SSCNC A (D SEARS) USASSCOM NRDEC BLDG 3 RM R-140 NATICK MA 01760-5015
1	ARL HRED FT HUACHUCA FLD ELEMENT ATTN AMSRL HR MY (B KNAPP) GREELY HALL (BLDG 61801 RM 2631) FORT HUACHUCA AZ 85613-5000
1	ARL HRED FT LEAVENWORTH FLD ELE ATTN AMSRL HR MP (D UNGVARSKY) TPIO ABCS 415 SHERMAN AVE RM 327 FT LEAVENWORTH KS 66027-1344
1	ARL HRED FLW FIELD ELEMENT ATTN AMSRL HR MZ (A DAVISON)* 320 ENGINEER LOOP STE 166 FT LEONARD WOOD MO 65473-8929
1	ARL HRED OPTEC FIELD ELEMENT ATTN AMSRL HR MR (D HEADLEY) PARK CENTER IV RM 1450 4501 FORD AVENUE ALEXANDRIA VA 22302-1458
1	ARL HRED SC&FG FIELD ELEMENT ATTN AMSRL HR MS (L BUCKALEW) SIGNAL TOWERS RM 207 FORT GORDON GA 30905-5233
1	ARL HRED STRICOM FIELD ELEMENT ATTN AMSRL HR MT (A GALBAVY) 12350 RESEARCH PARKWAY ORLANDO FL 32826-3276
1	ARL HRED TACOM FIELD ELEMENT ATTN AMSRL HR MU (M SINGAPORE) BLDG 200A 2ND FLOOR WARREN MI 48397-5000
1	ARL HRED USAFAS FIELD ELEMENT ATTN AMSRL HR MF (L PIERCE) BLDG 3040 RM 220 FORT SILL OK 73503-5600
1	ARL HRED USAIC FIELD ELEMENT ATTN AMSRL HR MW (E REDDEN) BLDG 4 ROOM 332 FT BENNING GA 31905-5400

NO. OF COPIES	ORGANIZATION
1	ARL HRED USASOC FIELD ELEMENT ATTN AMSRL HR MN (F MALKIN) HQ USASOC BLDG E2929 FORT BRAGG NC 28307-5000
1	US ARMY RSCH DEV STDZN GP-UK ATTN DR MICHAEL H STRUB PSC 802 BOX 15 FPO AE 09499-1500
	ABERDEEN PROVING GROUND
2	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CI LP (TECH LIB) BLDG 305 APG AA
1	LIBRARY ARL BLDG 459 APG-AA
1	ARL HRED ERDEC FIELD ELEMENT ATTN AMSRL HR MM (R MCMAHON) BLDG 459 APG-AA

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1998	Final

4. TITLE AND SUBTITLE

The Effects of Movement and Intravehicular Versus Intervehicular Communication on C^2V Crew Performance: Limited User Test Phase III

5. FUNDING NUMBERS

AMS: 622716.H700011
PR: 1L162716AH70
PE: 6.27.16
DAAL03-91-C-0034

6. AUTHOR(S)

Beck, H. P. (Appalachian State University); Pierce, L.G. (ARL)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Appalachian State University
Department of Psychology
Boone, North Carolina 28608

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
Human Research & Engineering Directorate
Aberdeen Proving Ground, MD 21005-5425

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARL-TR-1428

11. SUPPLEMENTARY NOTES

The contracting officer's representative (COR) is Dr. Linda Pierce, U.S. Army Research Laboratory, ATTN: AMSRL-HR-MF, Fort Sill, OK 73503-5600 (telephone 405-442-5051).

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The procedure and results of the group performance component of the command and control vehicle (C^2V) Limited User Test (LUT) Phase III are described in this report. The test was conducted to examine (a) the effects of movement on the ability of crews to work effectively as a team, (b) terrain impacts on team performance tasks, and (c) the effect of distributed team operations. Sixteen National Guardsmen, divided into four-person teams, served as participants. The evaluation design was similar to a 2 (Movement:Stationary, Moving) x 2 (Terrain:Paved, Course A) x 2 (Communication:Intravehicle, Intervehicle) with the baseline occupying the position of the nonfitting control arrangement. The effects of movement on team performance were evaluated by conducting some trials while the C^2V was stationary and other trials while it was moving. The influence of terrain on team performance was studied by conducting some trials on Course A of the Perryman test course and the remaining trials on a paved 3-mile course. In the intravehicular communication condition, the four members of a team were housed in the same C^2V and worked together on the same task. Teammates had visual contact and communicated verbally via intercom. Two teammates were in each C^2V for the intervehicular manipulation. It was concluded that the C^2V environment impaired all group performance tasks, especially those that appeared to demand a great degree of coordination and integration. Team performance was below the baseline when crews were housed in the C^2V, regardless of whether the vehicle was stationary or moving, although movement increased the deleterious impact of the C^2V on group performance. The impact of terrain on performance was inconclusive, possibly because of the small sample size and the limited number of situational conditions examined. If the C^2V is to become a prominent part of the 21st century Army's arsenal, then additional experimentation must be conducted to assess implications for team performance during a variety of conditions using validated task procedures.

14. SUBJECT TERMS

C^2V	teams
performance measurement	test and evaluation

15. NUMBER OF PAGES 45

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102